A mathematical model that predicts human biological age from physiological traits identifies environmental and genetic factors that influence aging

Authors: Sergiy Libert[1*], Alex Chekholko[1], Cynthia Kenyon[1*]

1, Calico Life Sciences, South San Francisco, CA, 94080
*cynthia@calicolabs.com
*libert@calicolabs.com

Keywords

Human aging, UKBB, longevity, GWAS, healthspan, biological age, physiological clock, cognition

Abstract

Why people age at different rates is a fundamental, unsolved problem in biology. We created a model that predicts an individual's age from physiological traits that change with age in the large UK Biobank dataset, such as blood pressure, lung function, strength and stimulus-reaction time. The model best predicted a person's age when it heavily-weighted traits that together query multiple organ systems, arguing that most or all physiological systems (lung, heart, brain, etc.) contribute to the global phenotype of chronological age. Differences between calculated "biological" age and chronological age (ΔAge) appear to reflect an individual's relative youthfulness, as people predicted to be young for their age had a lower subsequent mortality rate and a higher parental age at death, even though no mortality data were used to calculate ΔAge. Remarkably, the effect of each year of physiological ΔAge on Gompertz mortality risk was equivalent to that of one chronological year. A Genome-Wide Association Study (GWAS) of ΔAge, and analysis of environmental factors associated with ΔAge identified known as well as new factors that may influence human aging, including genes involved in synapse biology and a tendency to play computer games. We identify a small number of readily measured physiological traits that together assess a person's biological age and may be used clinically to evaluate therapeutics designed to slow aging and extend healthy life.

Introduction

The process of aging is universally similar yet deeply unique to each person. By observing a person for a moment, one can deduce their age with high accuracy, even though no two people age the same way. Some individuals might lose hair with age or develop chronic diseases, whereas others might not. Investigating both the universal aspects of aging as well as the basis of individual differences, and developing means of measuring physiological age and health, will provide opportunities to improve human lives.

The rate of aging; that is, the rate at which organisms lose physiological fitness and accumulate morbidity, has both genetic and environmental determinants. Humans age more slowly than, for example, dogs, so genes play a key role, but environmental factors like smoking and exercise influence aging as well. In this study, we have used publicly available data of human health parameters to systematically identify genetic and environmental variables that influence human aging.

To generate an inclusive, wholistic model of human aging, we queried a large, well-annotated human database (the United Kingdom BioBank or UKBB) comprising over 3000 phenotypes that together span the functions of multiple organs and physiological systems. The UKBB's medical, environmental, and genetic data on ~500,000 British volunteers is a unique resource to investigate the biology of aging. While participants of UKBB are not a random cross-section of society (Abdellaoui et al., n.d.; Haworth et al., 2019), this rich database nonetheless likely provides generalizable insights into human aging and disease (Hanlon Id et al., 2022).

A number of published studies describe and employ methods to identify genes that might influence human aging. The majority of those studies (Graham Ruby et al., 2018; Pilling et al., 2017; Wright et al., 2019) focus on lifespan (Joshi, n.d.; Timmers et al., 2022, 2019); for example, age at death or parents' age at death, or analyze cohorts of people with exceptional lifespan (Bae et al., 2022; Shen et al., 2020); or the presence or absence of one or few age-associated diseases (Timmers et al., 2022, 2020; Zenin et al., n.d.). Additionally, researchers have used molecular traits, such as blood proteins (Coenen et al., 2023), or blood DNA methylation patterns to build and analyze biological age prediction algorithms (clocks) to identify genes that influence aspects of human aging (Gibson et al., 2019; Lu et al., 2018; McCartney et al., 2021). Biological age clocks derived from one or few physiological measures have also been constructed, such as a biological clock built using 3D facial scans (Xia et al., 2020). Likewise, a biological clock built using the gut microbiome (Wilmanski et al., 2021) was used to identify individuals who might be aging slower or faster than average and suggest drugs that might influence gut health. Recently, aging of separate organs has been investigated and linked to age-associated diseases and mortality (Tian et al., 2023), and biological age has been estimated using AI methods (Qiu et al., 2022).

In our approach, we set out to measure human aging directly and wholistically, making sure that all systems relevant to health are represented. To do so, we sampled and analyzed traits reporting on multiple organ systems and physiological domains. We quantified the markers of aging that reflect overall physiological health, such as strength, stimulus-reaction time and blood pressure. This multi-systemic approach does not rely on a presence or absence of recognized diseases or a small number of binary events, such as death or stroke, and therefore reflects human aging more directly. Likewise, instead of concentrating on diseases, we aimed to evaluate a multitude of physiological parameters that change in "healthy" people, to allow us to identify factors missed by previous studies.

We developed a series of mathematical models that consider 121 age-related traits and predict a biological age for each individual. We show that the model that best predicts age incorporates data reflecting the activity of most if not all the organs and physiological systems. By comparing predicted biological age to actual age, we identified individuals who may be aging slower or faster than average. Using this model, we identified new environmental factors and genetic loci that may influence biological age. By building models lacking clusters of phenotypically correlated (typically organ-specific) traits, we further categorized these genetic

loci and environmental factors as those likely to influence aging globally vs those that likely impact a single organ system. Likewise, by analyzing a smaller, healthier sub-cohort of UKBB participants, we identified factors likely to influence apparent age by conferring an age-related disease. Notably, our findings highlighted neural function as an important determinant of overall biological age. Finally, after analyzing the performance of different physiological clocks, we identified twelve key physiological traits that together could measure biological age in longitudinal clinical trials for interventions that increase human healthspan.

Results

Physiological traits that change with age

To identify age-dependent traits, we conducted linear regression analysis on every UKBB parameter relative to the age of the participants (see Supplementary Table 1a, b) and recorded the list of those with a non-zero slope and adjusted statistical significance better than 10^{-3} (see Supplementary Table 2a, b). Examples include systolic blood pressure (shown in Fig. 1a), which increases with age, and hand-grip strength (shown in Fig. 1b), which decreases with age.

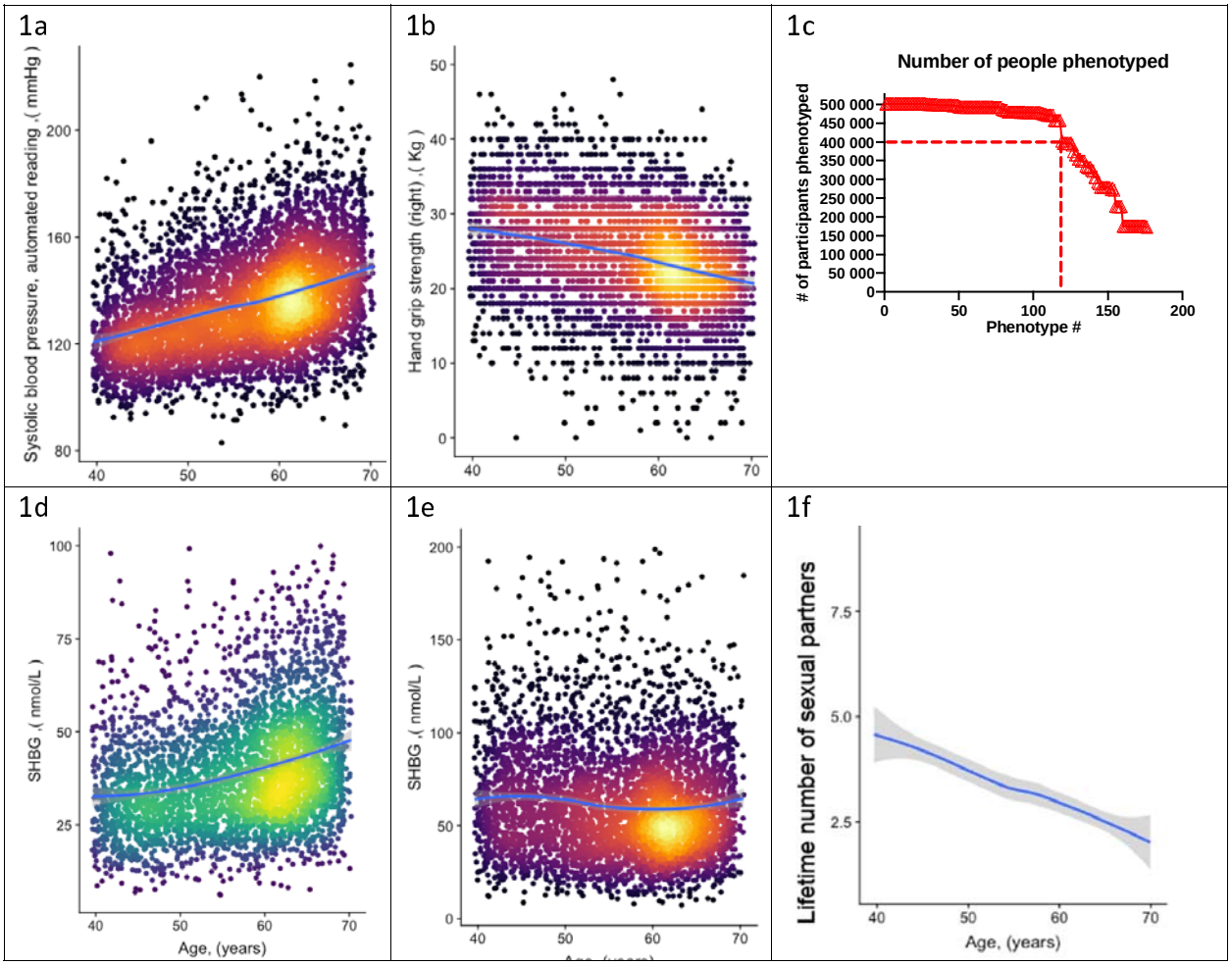


Figure 1. a) Systolic blood pressure (UKBB field ID# 4080) and b) hand-grip strength (UKBB field ID# 47) of a random set of 10,000 female UKBB participants is plotted against their age; c) Number of age-sensitive phenotypes plotted against the declining number of people in whom these phenotypes were measured; d, e) sex hormone binding globulin concentrations (UKBB field# 30830) of a random set of 10,000 males (d) and females (e); f) Average number of lifetime sexual partners is plotted against the age of UKBB participants (UKBB field ID# 2149). Grey area denotes 99% confidence interval. Color of dots on the plot represents relative density of dots in the area.

Before proceeding to modelling, we needed to address three issues:

1. Certain phenotypes, such as MRI brain scans, were only available for a subset of UKBB participants (in this case <50,000). Therefore, we could not use MRI data to estimate the age of the remaining participants. Thus, the inclusion of such incomplete phenotypes in the UKBB database required an optimization strategy. The objective was to identify individuals who appeared young for their age, and the more individuals in the study, the greater the likelihood of discovering them. Likewise, including more diverse phenotypes improves the robustness and global assessment of overall aging. However, as we increased the number of age-dependent phenotypes, the number of individuals evaluated decreased. From the curve's shape (Fig. 1c), we estimated an optimal inclusion threshold to be ~120 +/- 15 phenotypes.

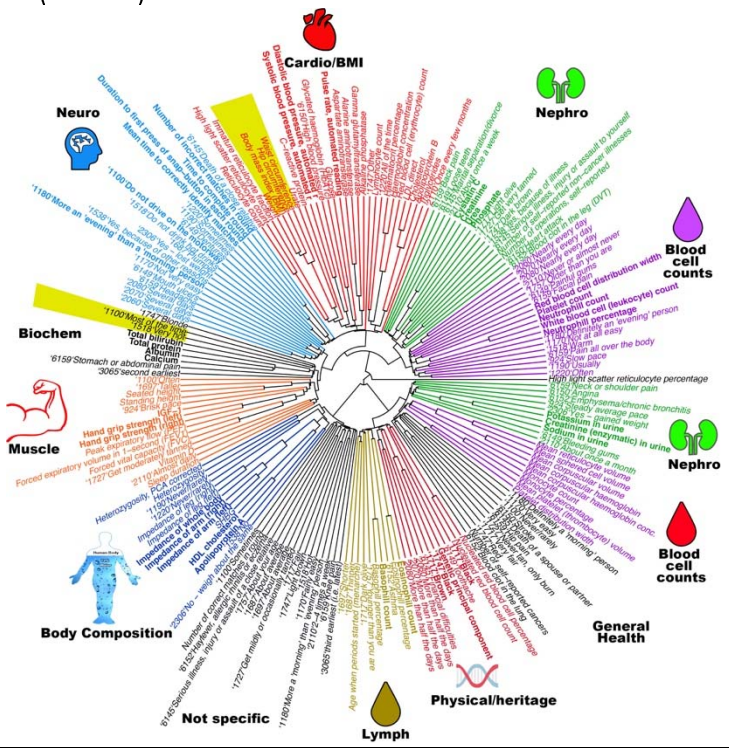
2. Significant phenotypic differences exist between the sexes. For example, the parameters “age at which first facial hair appeared”, “age at menopause” and “degree of pattern balding” are gender specific. Additionally, shared phenotypes may have different dynamics in males versus females. For example, increasing plasma concentration of sex-hormone binding globulin (SHBG) is one of the best predictors of age in males (Fig. 1d); however, in females, its plasma concentration stays nearly constant or even tends to decrease (Fig. 1e). Thus, we analyzed male and female aging separately.

3. The dataset we used is largely cross-sectional, meaning that each data point represents a different person at a different age. Consequently, phenotypes that are used to predict age could be indicative of cultural and societal changes over time, rather than biological changes associated with aging. For instance, a good predictor of age (with a p-value < 10^{-52}) is the lifetime number of sexual partners (Fig. 1f). While sexual activity and fertility have been linked to human aging and longevity (Min et al., 2012), the correlation here is most likely driven by evolving social norms in Britain. Other examples of such traits include "how many siblings do you have" or "how long have you lived in your current house." Moreover, some biological measurements were derived using age as a parameter. For instance, BMR (Basal Metabolic Rate) is an outstanding age predictor (p-value < 10^{-255}). However, BMR was not measured directly; instead, it was computed using a formula that incorporates height, weight, gender, and age itself. Therefore, we examined each age-dependent parameter independently, aiming to satisfy three broad criteria: a) the trait should not reflect societal norms and structures; b) the trait should not be a function of elapsed time (e.g., how long have you been drinking green tea?); and c) the trait's value should not depend on a person's actual age. We endeavored to use purely biological and physiological parameters. Although it is possible that the selected phenotypes were still influenced to some degree by the birth cohort, these considerations should have reduced this effect. The complete list of age-related traits we selected, along with the reasons behind our choices, can be found in supplementary tables S2 c and d.

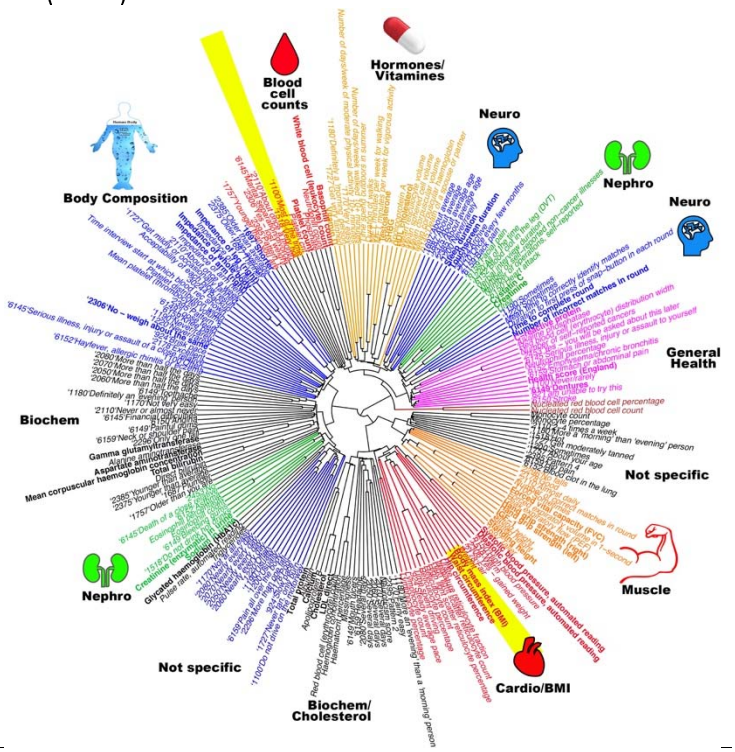
Age-dependent physiological traits fall into clusters

The phenotypes we selected for our age-prediction model were often correlated to one another; for example, left-hand and right-hand grip strength. To assess the degree and pattern of correlations among the age-dependent traits (see Supp. Table S1), we first normalized each phenotype by its mean and standard deviation. For phenotypes represented as multiple-choice questions (e.g., do you take naps - often, sometimes, rarely, never?), we encoded each answer option as a binary vector (one or zero), and these vectors were also normalized. Correlations were computed for each pair of phenotypes and visualized as dendrograms (fig. 2a, b). As expected, highly correlated phenotypes grouped together, such as "BMI"- "Weight"- "Waist Circumference" or "Cholesterol"- "LDL". Surprisingly, this analysis uncovered strong correlations that were not obvious, such as "I drive faster than the speed limit most of the time (id# 1100)" with "I like my drinks very hot (id# 1518)" (fig. 2a, b; marked with yellow shadows). Notably, most of the clusters appeared to be enriched for phenotypes associated with a specific organ or physiological system. For example, the cluster that contains "Creatinine", "Urea", "Cystatin-C" and "Phosphate" likely reflects kidney function; whereas the cluster that contains "Systolic blood pressure" and "Diastolic blood pressure" likely reflects cardiovascular function (fig. 2a, b). That said, upon close examination, it is not intuitively obvious why some physiological traits do or do not cluster with one another. Thus, this dendrogram might be a valuable data source for future hypothesis generation and exploration.

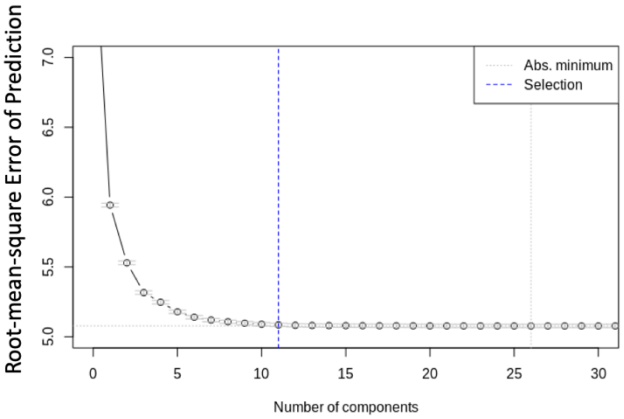
2a (female)



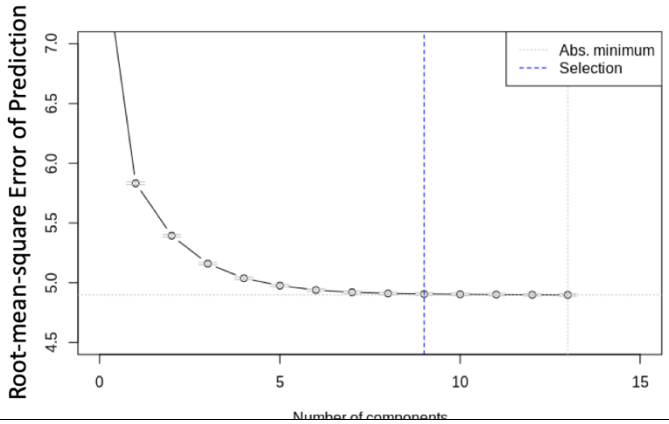
2b (male)



2c (female)



2d (male)



2e (female)

Parameter	Weight
Systolic blood pressure, automated reading	13.21
Forced expiratory volume in 1-second (FEV1)	9.25
IGF-1	8.82
Urea	8.09
Cystatin C	8.05
Mean time to correctly identify matches	6.63
Diastolic blood pressure, automated reading	5.99
Sodium in urine	5.45
Glycated haemoglobin (HbA1c)	5.28
Dentures	5.20
Alkaline phosphatase	5.14
Hand grip strength (left)	4.31
Calcium	4.17

2f (male)

Parameter	Weight
Systolic blood pressure, automated reading	12.22
SHBG	11.79
Forced expiratory volume in 1-second (FEV1)	11.26
Waist circumference	9.10
Cystatin C	7.84
Testosterone	7.33
Mean time to correctly identify matches	7.21
Dentures	6.77
Touchscreen duration	6.08
IGF-1	5.70
Body mass index (BMI)	5.64
Sodium in urine	5.55
Urea	5.43

2g (female)

2h (male)

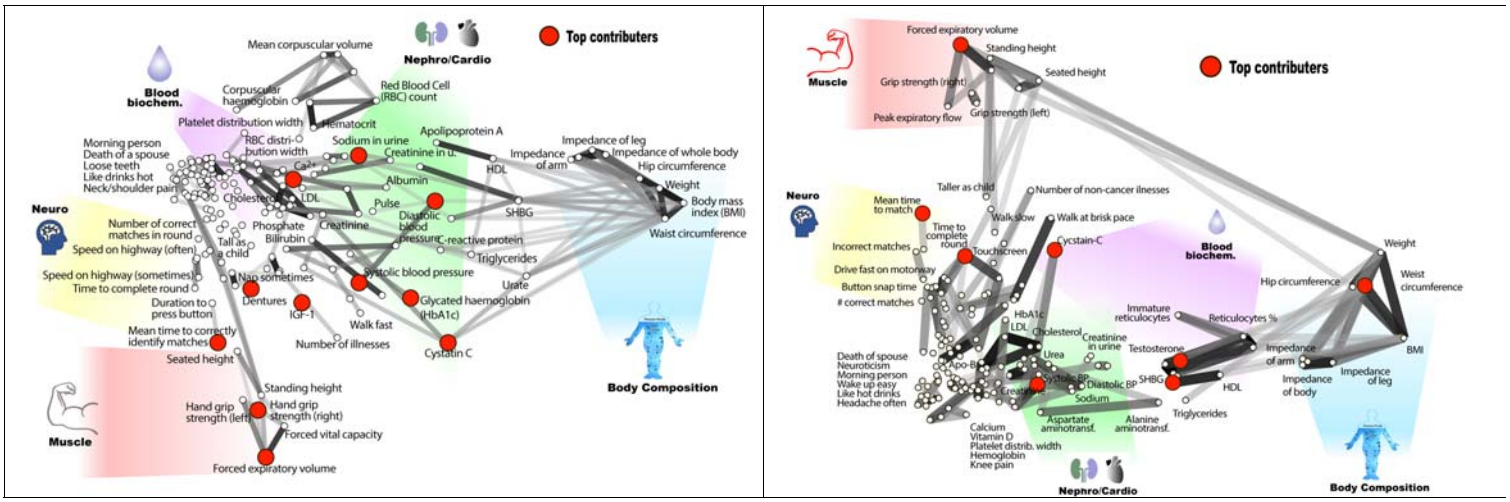


Figure 2. Age-dependent phenotypic clustering. Dendrogram plots of age-dependent female (a) and male (b) phenotypes selected for age prediction. Numbers in the name of "rays" represent UKBB ID numbers for multichoice questions (see supplementary table 1 or the UKBB website), followed by the answer. Major clusters were colored and subjectively assigned a name that reflects a possible biological function of the cluster; The number of principle components included in the PLS (Projection to Latent Structures) model to predict age vs root mean square error of the predictions is plotted for females (c) and (d) males; e) the top phenotypes with the highest weights in the age-predicting PLS model are listed for (e) females and (f) males. Phenotypes shaded in green are shared between sexes, red are specific to females, and blue are specific to males. All phenotypes were used for both sexes, and this shading reflects only the position in list of top-13 traits; g) list of phenotypes used to predict age of females and (h) males projected on 2D space using correlation as the distance measure. The degree of correlation is also depicted by grey lines, the darker the shade, the stronger the correlation. Note that the distortion in positioning is an inevitable consequence of projecting high-dimensional data into 2D space. As before, groups of related phenotypes were subjectively assigned a name that likely depicts their physiology, and phenotypes with the highest weight in the PLS model were depicted by red dots.

A mathematical model to predict age

To develop a model that predicts age, we experimented with several algorithms, including simple linear regression, Gradient Boosting Machine (GBM) and Partial Least Squares regression (PLS). The outcomes of these approaches were almost identical, likely due to the small number of predictors (121 phenotypes) and comparatively large number of participants (over 400,000). We selected PLS regression because it enabled us to determine the number and composition of components required to predict age optimally from the data.

PLS modeling is not tolerant of missing values, and in the UKBB dataset we used, over 60,000 participants (~15%) lacked at least one phenotypic measurement. To prevent excessive imputation, we excluded any individual missing more than 15 datapoints from the study, thereby decreasing number of selected female participants from 222,111 to 215,949 (~2.7% loss), and males from 188,609 to 183,715 (~2.6% loss). We imputed and scaled the values of the remaining participants with missing data (Methods).

Next, we determined how many PLS components (each derived from UKBB phenotypes) were required to predict chronological age. To do so, we constructed a series of age-prediction models using an increasing number of these components. The first model was built using only component #1, the second using components #1 and #2, and so on. At each step, we calculated the root-mean-square error of the age prediction and determined its decline using the R function "selectNcomp" (see Fig. 2c, d). Our analysis revealed that only 11 independent components were required to describe female aging dynamics, and 9 independent components were required for males. Including additional components did not further improve the model performance. Therefore, we used the R function "plsR" with 9 and 11 components for males and females, respectively, along with the Cross-Validation function (CV) to prevent overfitting when building models to predict age using UKBB phenotypes.

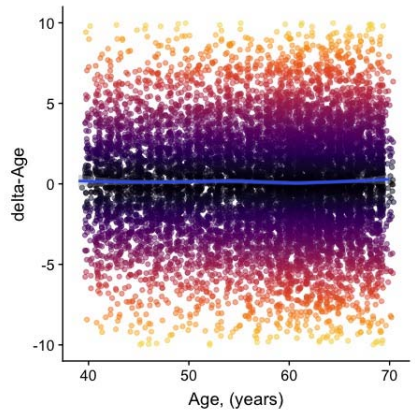
It was interesting to determine which individual age-sensitive phenotypes were most useful for age prediction. Since many phenotypes contribute to multiple PLS components, we deconstructed each PLS component and calculated the sum of the absolute values for phenotype coefficients across all components. This provided a weight metric for each phenotype used to predict age. The top thirteen phenotypes with the highest weights are presented in figures 2e and 2f. Most were shared between males and females and were associated with different physiological systems; for example, systolic blood pressure (which likely correlates with cardiovascular health), forced expiratory volume (pulmonary and cartilage/bone health), urea and cystatin C levels (kidney health), and mean time to correctly identify matches (cognitive health). Moreover, if we deleted one of these selected traits, the model substituted a close correlate; specifically, it substituted 1-second FEV (forced expiratory volume) for FEV, systolic blood pressure for diastolic blood pressure, and hand-grip strength (right) for hand grip strength (left). The fact that the model best predicted chronological age when it received input from a wide range of physiological systems underscores the global, systemic nature of the aging process. Similar conclusions were drawn from high-dimensional analysis of aging mice (Chen et al., 2022).

Inferred (biological) age predicts all-cause mortality better than chronological age

We utilized the physiological phenotypes listed in tables S2a, b and the PLS modelling described above to predict female age with a root mean square error of 4.8 years, R^2~0.63, and predict male age with a root mean square error of 5.1 years, R^2~0.6. Several factors may contribute to discrepancies between predicted biological age and chronological age, including statistical noise, variations in life histories among UKBB participants, limited accuracy of certain measurements, and inadequate numbers of relevant measurements. However, some of this discrepancy may arise because certain individuals are aging more slowly or rapidly than the mean for that age. Consistent with this interpretation, we observed a significant correlation of residuals between two assessments for a small number of UKBB participants who were evaluated longitudinally (twice) with intervals of up to 12 years (R2~0.56, $p<10^{-255}$).

To estimate biological age from this cross-sectional data, we computed a value termed ΔAge for each participant. We define ΔAge as the individual's chronological age subtracted from their predicted age and normalized such that the average ΔAge for the entire population at each age is zero. ΔAge is negative if an individual is predicted to be younger than they are and positive if an individual is predicted to be older. The ΔAge parameter carries no information about the person's actual chronological age, as it is equally distributed across zero at any age (fig. 3a). Comparable approaches have been employed previously, such as using DNA methylation patterns (Marioni et al., 2015), or facial images and computer vision (Chen et al., 2015) to predict age and identify potentially "fast agers" and "slow agers".

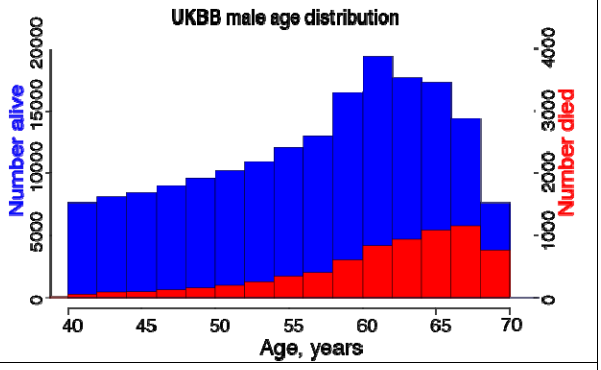
3a



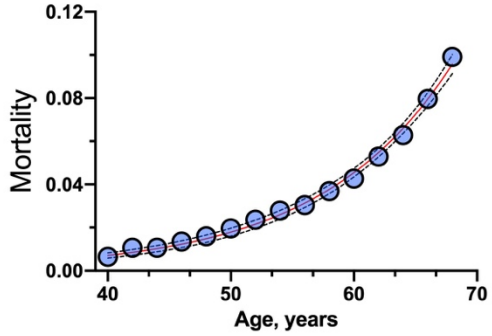
3b

red-died

blue-alive

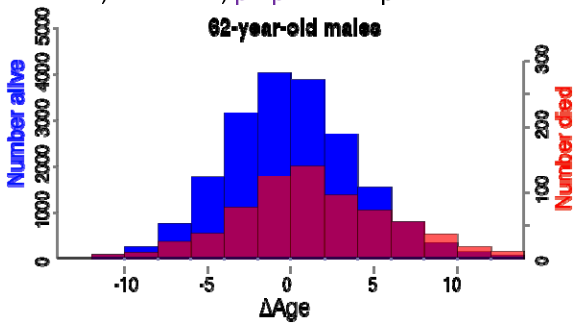


3c

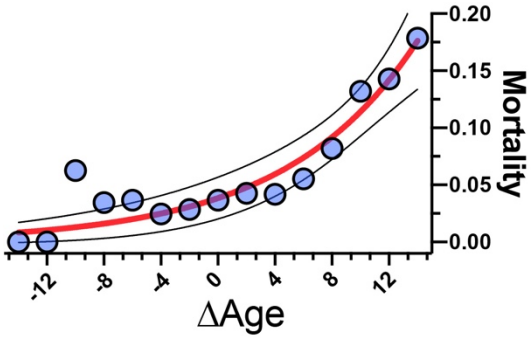


3d

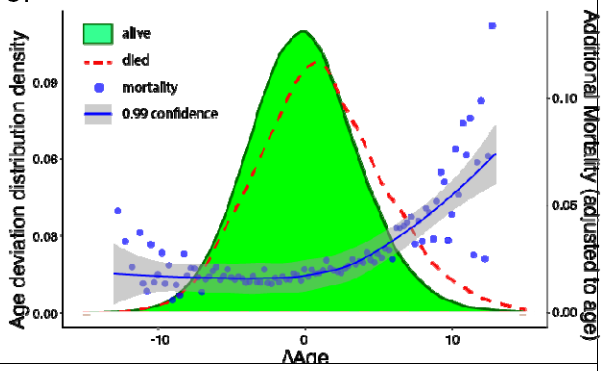
Red-died; blue-alive; purple-overlap



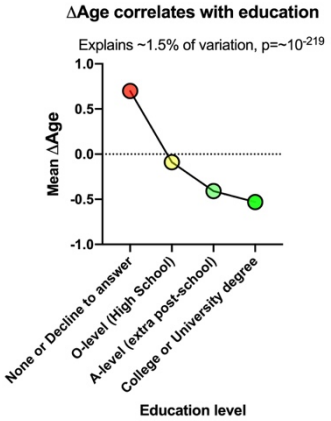
3e



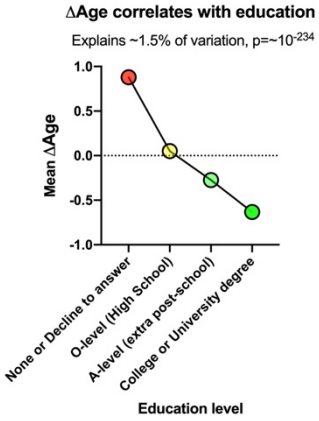
3f



3g (male)



3h (female)



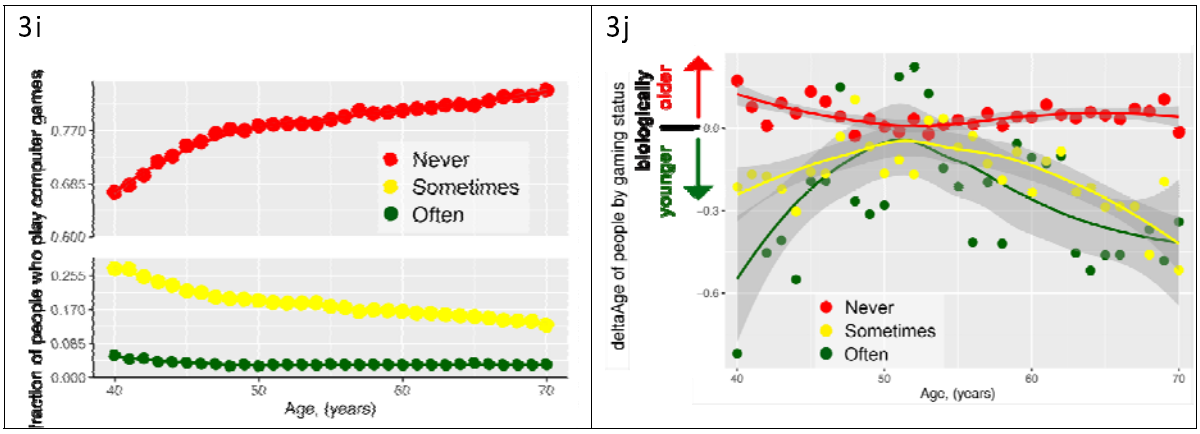


Figure 3. ΔAge has biological meaning. a) delta-age (ΔAge, predicted biological age minus chronological age) is plotted against chronological age for a random subset of 10,000 UKBB participants. Note that there is no correlation between age and ΔAge; b) histogram of age distribution (blue), and death distribution (red, right y-axis) is presented for UKBB males; c) mortality of UKBB male participants vs their age is platted, note the classical exponential (Gompertzian) shape. Blue dots are actual data, the red line is an exponential fit, and the black dash line is 95% confidence interval; d) histogram of the ΔAge distribution (blue), and death distribution (red, right y-axis) is presented for UKBB males of 62 years of age only; e) mortality of 62-year-old males is plotted against their ΔAge. Blue dots are actual data, the red line is an exponential fit, and the black dashed line is 95% confidence interval. Once again, note the classical exponential (Gompertzian) shape with ΔAge, even though all the subjects are the same age chronologically; f) distribution of ΔAge for all the people in UKBB (all ages and all genders, green shape). The distribution of ΔAge for people who died within 5 years after enrolling in the UKBB (red line) is shown for comparison; note a shift of the deceased distribution to the right, towards larger ΔAge (predicted older on average). The mortality penalty due to ΔAge is plotted as blue dots (left y-axis), the exponential fit of these data is presented as a blue line, and the 99% confidence interval as a grey shade; g) average ΔAge is plotted for UKBB males (g) and females (h) against their highest education (qualification) level achieved; i) the fraction of people who play computer games "sometimes" (yellow dots), never (red dots), and people who play computer games "often" (green dots); j) average Δage of people at different ages separated by their computer gaming habits (see 3i). As a group, people who play computer games "often" are biologically younger than people who play computer games "sometimes", or "never".

One year of ΔAge carries approximately the same mortality risk as one year of chronological age

The classical paradigm of aging described by Gompertz stipulates that mortality rates increase exponentially with time, doubling roughly every 8 years (Kirkwood, 2015). In the UKBB dataset that we analyzed, a small number of participants (8,883 males and 5,668 females, fig.3b) passed away within 5 years of their initial test-center attendance. The distribution of these deaths among UKBB participants has a typical "Gompertzian" shape, with mortality rates exponentially doubling every 7.7 years for both males and females (figure 3c). In Gompertz' model, where mortality depends only on age, everyone of the same age has an equal likelihood of dying. However, by incorporating ΔAge, we were able to further forecast death among individuals of the same age. To illustrate this point, consider males who are 62 years old and group them based on their ΔAge (as shown in figure 3d). Individuals on the left side (with negative ΔAge values) were predicted by the model to be younger than 62, while those on the right were predicted to be older. In this UKBB sub-cohort, several hundred subjects died within five years following their enrollment. Plotting the average mortality for each ΔAge bin in this stratification of 62-year-olds resulted in a Gompertz-like mortality distribution. Notably, the effect of one year of ΔAge on the mortality rate was almost identical to that of one year of chronological age. It is important to emphasize that death data were not considered during the development of the model of biological age or derivation of ΔAge, and that ΔAge does not exhibit any correlation with chronological age (as illustrated in fig. 3a). The capacity of ΔAge to predict mortality with a similar level of accuracy as chronological age is consistent across genders and ages and can even be observed when individuals of all ages are combined (fig. 3f). We consider this progressive increase in mortality rates with progressively larger ΔAge to be a powerful validation of this modeling strategy for assessing biological age. The fact that combining chronological age with ΔAge leads to a more precise prediction of mortality risk than relying on chronological age alone might be of interest to actuaries.

ΔAge correlates with parental lifespan

Remarkably, we observed a robust correlation of ΔAge with the age at death of the participant's father (p-value = $1.9*10^{-43}$ for females, and p-value = $3.9*10^{-31}$ for males), and mother (p-value = $1.1*10^{-68}$ for females, and p-value = $1.3*10^{-32}$ for males). Individuals predicted to be biologically younger had parents who lived longer. Previous studies have reported that the lifespans of parents and offspring are correlated (Graham Ruby et al., 2018; Milman and Barzilai, 2016). These findings, too, provide strong validation for the model, reinforcing the idea that ΔAge is not simply noise, but rather carries significant information about the aging process and its variability in the population.

Environmental factors that influence biological age

Previous studies have shown that personal wealth is positively associated with human lifespan (Chetty et al., 2016; Wang and Geng, 2019), whereas smoking and excessive drinking are negatively associated with lifespan. To investigate whether this measure of physiological ΔAge has similar associations, and possibly to identify new environmental factors that influence aging, we calculated the correlation of ΔAge with every parameter available from UKBB (supplemental tables S3a, b). Correlations with p-values lower than 10^{-5} (calculated to correct for multiple testing) were considered statistically significant. Interestingly, we observed a strong association of ΔAge with age-dependent biological phenotypes that were not included in the model to predict ΔAge due to the low number of people who underwent the assessment. For example, heel bone density (UKBB field #3148) and Thalamus volume (UKBB field #25011) both had strong associations with ΔAge (p-values were ~10^{-11} and 10^{-10}, respectively). These and other phenotypes with strong ΔAge correlations again help to validate the model and might be useful parameters to consider when building biological clocks in the future.

Tables S3c and S3d list the environmental factors we found to correlate with ΔAge. As predicted, wealth was positively correlated with a more youthful ΔAge. For instance, parameters such as "home location" (UKBB field id# 20075), "place of birth" (UKBB field id# 129), "Townsend deprivation index" (UKBB field id# 189), and "total income" have a strong and significant correlation with ΔAge (Tables S3). Additionally, smoking and exposure to smoke (UKBB field ids# 20161 and 20162) were positively correlated with an older ΔAge. The impact of moderate alcohol drinking on long-term health is still a subject of debate. In our data, the overall frequency of alcohol consumption (numerous UKBB fields, like 20414) did not have a significant correlation with ΔAge, however, the alcohol type did. Consuming beer and hard cider (UKBB field id# 1588) were positively correlated with ΔAge, whereas consuming Champagne and other white wines (UKBB field id# 4418) was negatively correlated. It is likely that drinking Champagne frequently is an indicator of higher socio-economic status.

The single most significant non-biological parameter that correlated with ΔAge in both males and females (p-value<10^{-200}) was "Qualifications" or the level of education achieved (UKBB field id# 6138). Each additional level of education was progressively associated with increased "youthfulness" (Fig. 3g, h). Interestingly, the effect size of education (-1.51) was much greater than that of wealth (-0.81) or place of birth (-0.13).

Certain leisure and social activities were also correlated with ΔAge. The amount of TV watching (UKBB filed# 1070) was positively correlated with ΔAge in both males and females, whereas time spent outdoors (UKBB filed# 1050) for males, and DIY projects (UKBB filed# 2624) for females were correlated with younger ΔAge. Intriguingly, the second strongest behavioral trait that associated with ΔAge was the “frequency with which people play computer games”. This is a novel association, and one that is less likely to reflect socioeconomic status, as access to computer gaming is inexpensive and widely available. Playing computer games associated with youthfulness (fig. 3i, j, Supp. Item #1), with a size effect of -2.2 and p-value of $4*10^{-8}$. This association was equally strong if “age” was factored out from the regression, indicating that generational changes in leisure activities do not explain this association.

Genetic loci associated with biological age

To identify potential genetic determinants of physiological ΔAge, we carried out a genome-wide association study (GWAS), using linear models separately on males and females (Methods). Manhattan plots for male and female GWAS models are presented in figures 4a-d (for summary statistics, see supplementary tables S4). The inflation factor in our analysis was λ_{gc}=1.2005 for males and λ_{gc} =1.2531 for females. Linkage disequilibrium regression intercepts were 1.0213+/-0.0083 and 1.0285+/-0.0119 for males and females respectively.

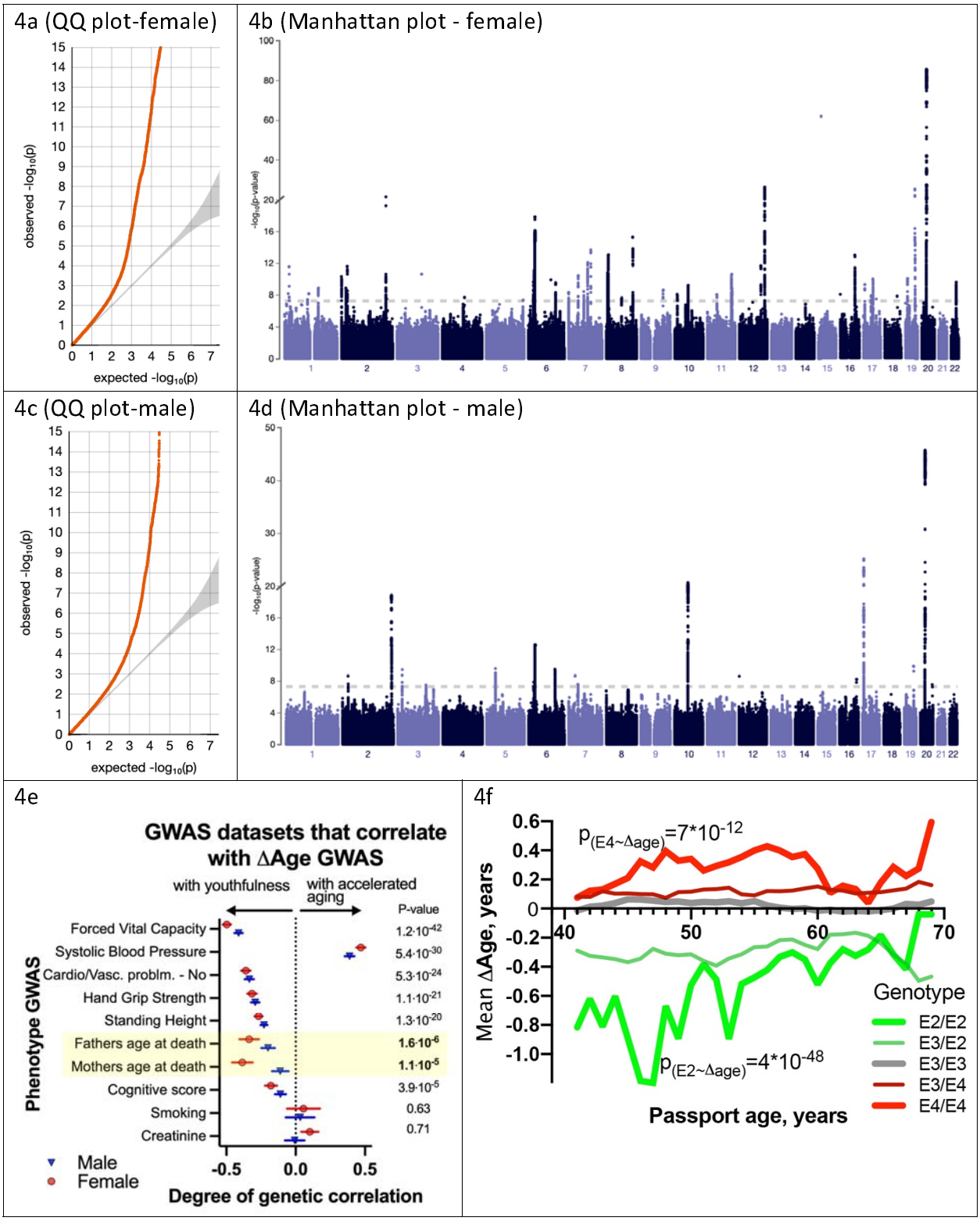


Figure 4. Genetic analysis of ΔAge. a, c) Quantile-Quantile plots for female and male -log10 p-values; b, d) Manhattan plots from genome-wide association analysis of female and male ΔAge; e) Correlation of ΔAge GWAS determination with other GWAS performed and reported by the UKBB consortium. Note the strong genetic relation between GWASs for ΔAge and parental age at death; f) effect of *APOE* alleles on average ΔAge plotted across different ages. Beneficial *APOE2* alleles are in green, and detrimental *APOE4* alleles are in red.

Using a stringent multiple testing correction for GWAS (Chen et al., 2021) with a threshold of 10^{-9}, we identified 9 loci associated with ΔAge in males and 25 loci in females (fig. 4a, b, table S4). Four of these loci were found in both sexes. Specifically, these include the HLA locus, located at chr6:32,600,000; chr10:64,900,000, a locus that contains NRBF2, JMJD1C, and TATDN1P1 genes; chr19:45,413,233, a locus that contains APOE, TOMM40, and APOC genes; and chr20:23,613,000, a locus that contains the CST3 gene. These genes are strong candidates to influence whether a person is biologically young or old for their age. Two of these loci, APOE (Schächter et al., 1994; Sebastiani et al., 2019) and HLA (Yang et al., 2017), have previously been associated with human longevity, which increases our confidence in the analysis. GWAS analysis of combined male and female ΔAge data identified 12 additional loci (and candidate genes associated with these loci), are listed in table S4, and figure 5b.

A healthy sub-cohort distinguishes genes that affect aging vs age-related disease

Some genes that associated with ΔAge in our analysis are known disease risk factors. For example, the HNF1A (hepatocyte nuclear factor 1 homeobox A) locus (top SNP – rs1169284, ΔAge association p-value = $3.0*10^{-23}$) is associated with diabetes (Shepherd et al., 2009) and cancer (Abel et al., 2018). The APOE (apolipoprotein E) locus (top SNP – rs7412, ΔAge association p-value = $4.4*10^{-33}$) is associated with Alzheimer's disease and coronary heart disease (Xu et al., 2016).

It is possible that people who carry risk alleles for age-related disease have a higher ΔAge due to the disease itself, even though their aging may be unaffected otherwise. To investigate this, we calculated the association of top loci with ΔAge in a "healthy-only" cohort, excluding people who had been diagnosed with disease; specifically, diabetes, cancer, asthma, emphysema, bronchitis, chronic obstructive pulmonary disease (COPD), cystic fibrosis, sarcoidosis, pulmonary fibrosis tuberculosis, any vascular or heart problems (such as high blood pressure, stroke, angina, or heart attack) or anybody with a history of allergic complications. These exclusion criteria decreased the number of the people in the study by almost 50%, however, the association of ΔAge for top hits remained (supplementary table 4). These findings suggest that most of the genetic signal associated with ΔAge comes not from a few susceptibility alleles for specific diseases but rather from alleles that describe and possibly drive fundamental processes that change with age; that is, possibly with aging itself. Conversely, this analysis also identified genes that were specifically responsible for certain diseases that present similarly to accelerated aging. For instance, the GCKR (glucokinase regulatory protein) locus showed a strong association with ΔAge (p-value=$8*10^{-12}$); however, the association disappeared when we excluded individuals diagnosed with diabetes. This demonstrates that mutations in GCKR cause a disease that resembles aging but do not have a detectable effect on the overall aging of healthy individuals.

Nonetheless, caution should be exercised when interpreting the analysis of this smaller, "healthier" subpopulation. It is possible that certain hits disappeared not due to disease but because of decreased statistical power resulting in false negatives. Conversely, some individuals may have had undiagnosed or subclinical disease, leading to false positives. Additionally, some of the associations may be false positives due to Collider bias. Thus, we favor the interpretation that among the GWAS hits that disappeared in the healthy sub-cohort were disease-susceptibility genes, while those that persisted likely influence the aging process more

generally. Future longitudinal and other studies in humans and potentially animals could lend support to this interpretation.

Heritability of ΔAge

To estimate heritability, we performed Linkage Disequilibrium (LD) score regression analysis (Zheng et al., 2017). The analysis involved 1,293,150 unique SNPs with an allele frequency higher than 0.01. We found that total genetic heritability (H^2) of ΔAge was ~11% (0.108+/-0.009) for females and ~10% (0.096+/-0.008) for males, which is similar to the genetic heritability estimated previously for human longevity (Graham Ruby et al., 2018; Melzer et al., 2020). This may be because the variation in genetic diversity is not substantial or because existing alleles of critical longevity genes do not have significant effect sizes in this human population.

GWAS signatures that correlate with the ΔAge GWAS

Another way to infer the biological meaning of ΔAge is to compare the GWAS signatures (Manhattan Plots) of ΔAge to GWAS signatures of other traits in public databases (Zheng et al., 2017). We found that the genetic signatures of some of the components used to calculate ΔAge were correlated to the genetic signature of ΔAge itself (Fig. 4e). For example, GWAS of Forced Vital Capacity (FVC) had a correlation with ΔAge GWAS of 0.49 +/- 0.02 (p-value=$5*10^{-65}$). In fact, remarkably, the most similar GWASs together spanned multiple organ systems (pulmonary, cardiovascular, musculature, cognition), arguing that this “aging” GWAS integrates the health of multiple organ systems.

In contrast, GWAS signatures of certain physiological parameters, such as blood creatinine levels, which were explicitly used in ΔAge derivation, had no genetic correlation with ΔAge (0.1+/-0.07, p-value=0.1). It is possible that traits whose GWAS signatures genetically correlate with the GWAS signature of ΔAge are drivers of aging, while traits with uncorrelated GWAS signatures are simply biomarkers. Certain metabolic parameters have been correlated to mortality in previous studies (Deelen et al., 2019), but it has been an open question if those metabolites have causal relationship to aging and mortality.

It is interesting to note that the genetic signature of ΔAge has a strong similarity to the genetic signature obtained through GWAS for "Father's age at death" and "Mother's age at death" (fig. 4e). This correlation was present even though the mortalities of subjects or parents were not part of the model and were not considered throughout the analysis. The genetic correlation of GWAS for parent's age mortality with GWAS for offspring's ΔAge was 0.39±0.03, p-value=$1*10^{-7}$ for females and 0.2±0.05, p-value=$3*10^{-5}$ for males. These GWAS correlations further demonstrate that ΔAge carries information about aging and longevity, despite its values being derived from cross-sectional physiological data and being independent of lifespan.

Gene ontology highlights a neuronal influence on biological age

To investigate whether specific pathways or systems have an influence on biological age, we performed GeneOntology analysis of extended GWAS hits (combined male and female genetic loci identified by the closest ORF). Five enriched pathways were identified in this analysis (Supp Item #2). Unexpectedly, the top enriched category (GO:98815) was *modulation of excitatory postsynaptic potential*, enriched ~18 fold over the expected by-chance reference, with multiple-testing-adjusted p-value of 0.046. This category was exceptional (~18-fold enrichment), as the

second-best enrichment category was enriched only ~3 fold (*response to oxygen-containing compounds*). This GO category comprised multiple genes influencing synaptic function (Supp Item #2) suggesting that the nervous system plays a particularly important role in aging systemically. Like the vasculature, the sympathetic nervous system impacts the function of many peripheral organs, and synapse function plays a critical role in the function and the maintenance of the CNS. Hints of such an association have come from genetics studies of worms (Apfeld and Kenyon, 1999; Li et al., 2016), flies (Libert et al., 2007) and laboratory rodents (Garratt et al., 2022; Zullo et al., 2019). It is possible that synapse function accounts for the association of computer gaming with ΔAge.

Cluster-dropout analysis enriches for GWAS hits that influence aging globally

If a GWAS hit influences aging itself; reflecting the function of all the organs and physiological systems, the association between the SNP and ΔAge should not disappear if any one measurement is omitted from the model. Thus, we investigated the robustness of the GWAS hits in a systematic way, using what we term "Cluster Dropout Models". Specifically, we constructed a series of male and female models to predict ΔAge by systematically excluding small sets of highly correlated phenotypic clusters. We built 10 models, in which phenotypic clusters related to muscle (drop-out model 1), body composition (2), kidney health (3), cardio health (4), blood cell composition (5), blood biochemistry (6), neuro-psychiatric phenotypes (7), lipid metabolism (8), physical attributes (9), or general health (10) were excluded. The list of phenotypes belonging to each cluster is reported in Supplementary Table 2 and was guided by the clustering presented in Figures 2a, b, g, and h. As expected, the ΔAge values remained consistent among all the drop-out models (Figure 5a). This means that if a person was predicted to be ~x years younger or older than their chronological age, this prediction was approximately the same regardless of the phenotypic clusters omitted.

A systematic evaluation of Cluster-Dropout models can suggest which of the genetic hits from our original full-model GWAS are likely to influence organismal aging and which are linked to a narrower phenotype. To perform this analysis, we took the best SNP from each candidate GWAS locus from the full male or female analysis (above) and tested its association with ΔAge computed using each of the 10 drop-out models. The bubble-plots in figure 5b represent the effect size of each of these SNPs (via bubble size), and the associated p-value (via color).

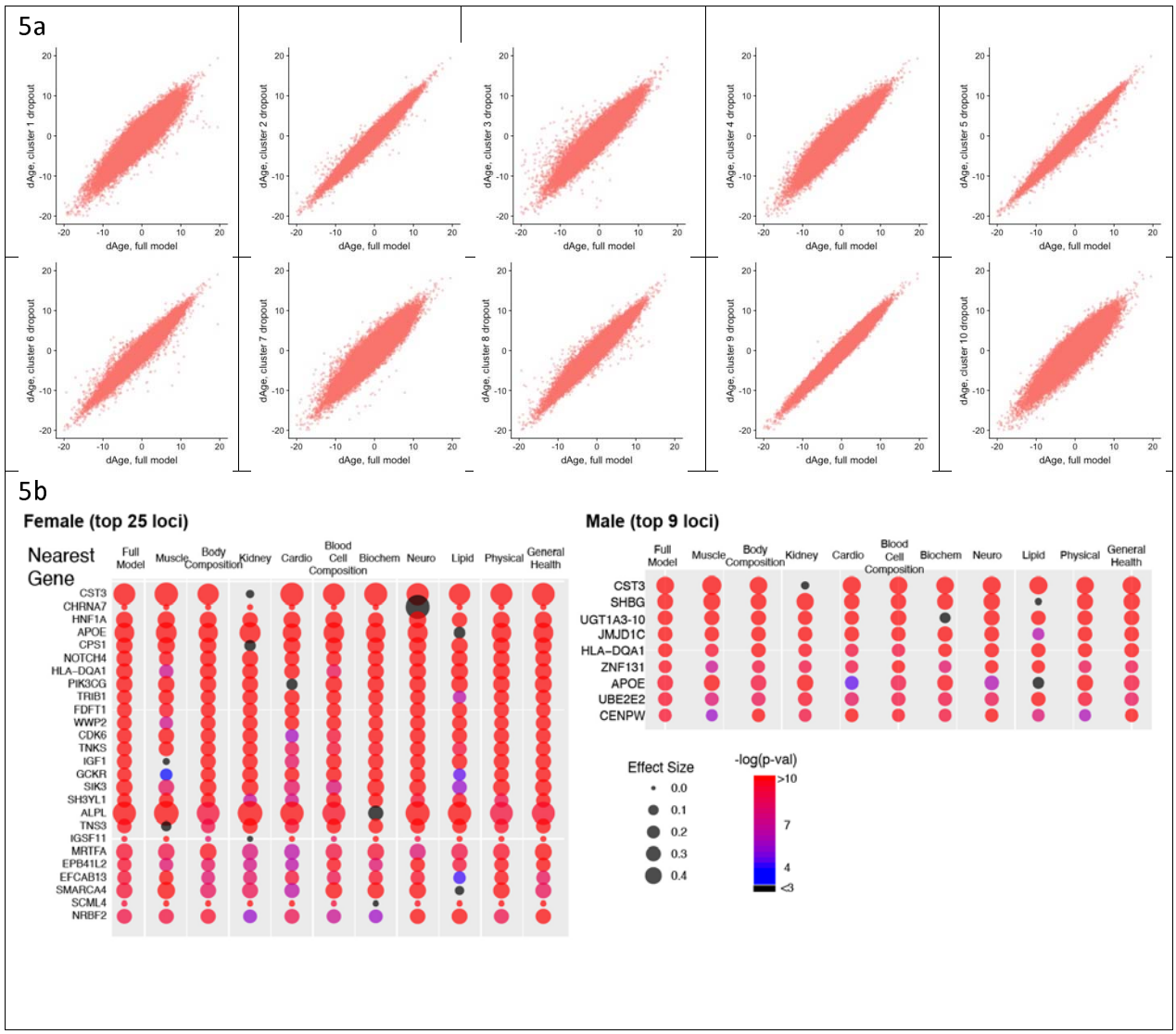


Figure 5. Cluster Dropout Models. a) Correlation between ΔAge calculated using full set of identified parameters and each of ten dropout models. Note that ΔAge values remain robust between models, meaning that if the person is predicted to have large ΔAge by the complete model, the "dropout" models will predict large ΔAge as well; b) The list of genes nearest to GWAS loci that associate with female and male ΔAge in the full model. Each hit is presented as a bubble, colored according to the significance of association of the locus with ΔAge, with size representing the effect size of the top SNP in the locus. The full summary is reported in Supplementary table 4.

As predicted, some GWAS hits disappeared in certain drop-out models. A particularly informative gene was CST3. CST3 encodes cystatin-C, a metabolite whose concentration increases with age. Levels of cystatin-C are routinely used to evaluate kidney health and it is proposed to be used as a marker in human aging study "TAME" (Justice et al., 2018). Elevated levels of this metabolite had been linked to elevated risk of cardiovascular disease (CVD) (van der Laan et al., 2016), risk of cancer (Jones et al., 2017), and neurodegeneration (Kaur and Levy, 2012). However, in a Mendelian Randomization Study (van der Laan et al., 2016) it was shown that while levels of cystatin-C predict CVD well, SNPs that robustly alter expression of cystatin-C do not associate with CVD.

In the full model, CST3 had the most significant association with ΔAge (effect size >0.4, p-value<10^{-80}) in both males and females, as represented by its large red bubble. This association remained significant in all the dropout models, except dropout number 3 (kidney health clusters), which contains the CST gene product, Cystatin-C concentration, which was one of the UKBB phenotypes used to generate the model. When the kidney clusters were omitted, the size effect of the CST3 association decreased to less than 0.1, p-value ~ 0.1, which is represented by the small black bubble. Likewise, if we calculated ΔAge using all the inputs in the full model except for "cystatin-C levels", the *CST3* locus was no longer associated with ΔAge. Combined, these data suggest that cystatin-C is a "marker" rather than a driver or determinant of aging. In contrast, some GWAS hits never dropped out, and these remained candidates for fundamental determinants of physiological ΔAge.

In the same way, cluster-dropout models can be used to interrogate environmental factors. For example, as described above, computer gaming correlates with a youthful biological age (Figures 3i,j, Supp. Item 1). The natural question is - are there specific physiological phenotypes, such as stimulus-reaction time or pattern recognition that drive this correlation or is it reflective of a "whole-body" biological age. To answer this question specifically, as well as to investigate all the phenotypes systematically, we calculated the strength of the correlation between every UKBB phenotype and all the cluster-dropout models (Fig. 5a) in both males and females (presented in supplementary table 5). To account for multiple testing, the Bonferroni corrected threshold of significance was $7*10^{-7}$. The correlation between biological age and computer gaming remained significant across all the models tested in both males and females, suggesting that there are no specific singular phenotypes responsible for this correlation. Such robustness of association was true for most phenotypes, but not all. For example, particulate air pollution (pm10) is associated with older biological age (p-value=$1.6*10^{-9}$ for females), however, if the model omits the cluster containing lung parameters, such as FEV, the correlation drops below Bonferroni-corrected statistical significance (p-value=$5*10^{-3}$ for females). This might suggest that particulate pollution mostly affects pulmonary health and to a lesser extent global organismal aging. One must keep in mind the caveats and complexity of comparing correlations of different phenotypes to each other, yet this dataset provides a good starting point for possible investigations of environmental factors influencing human aging.

DISCUSSION

General Caveats

Our study has several caveats. We used a cross-sectional dataset, where different ages are presented by people born at different times. Therefore, there is a likely a ”cohort effect” in some or all predictors we use. Additionally, our model assumes that the rate of aging is constant for each individual, which is not always true. For example, a person's aging rate may change if they stop smoking. Despite these modelling assumptions, we believe that the final results are valid and generalizable and allow us to suggest new methods to measure physiological aging in humans and identify new targets to slow down human aging. The robustness of our modelling can be also assessed by considering a small number of UKBB participants (~13,000 out of ~500,000), who have been assessed twice, with the follow-up intervals ranging from 4 to 12 years. We observed a significant correlation (R^2~0.6, p-value<10^{-255}) between biological-chronological age measures for these individuals between their two assessments. This suggest that variation due to noise is not large. We also found that there is a significant correlation between longitudinally calculated rates of aging (change in biological age divided by assessment interval) and the values calculated using cross-sectional approach. Furthermore, to minimize the cohort effect in our genetic analysis, we used the year of birth as a covariate. Together with the correlations we observed between Δage and mortality, parent’s mortality, previous GWAS longevity hits, and GWAS Manhattan plot comparisons, these findings suggest that the method we describe is a feasible approach to measure an individual's rate of aging and to identify genetic and environmental factors that may influence it.

Broader implications of the model for physiological aging

How a general term like “aging” maps onto age-dependent physiological traits is a deep question that may never be answered with great precision. In general, biological clocks can be used to identify new genes and environmental factors that influence aging, as we did here using this physiological clock. In addition, one can “look into the clock” itself to gain additional insights. For example, we found that this mathematical model could best predict chronological age when all the different organ and physiological systems were sampled, emphasizing the systemic nature of aging. If the phenotypes associated with chronological age resulted from the decline of only one or a few organs, this would not be the case. Second, the model showed how different physiological traits co-vary and cluster in the population. Some correlations, such as vitamin D and sleep duration, are not immediately clear. However, a post-hoc examination of such an association can be explained in the light of previous medical research. We anticipate that exploring analogous non-intuitive clusters that cannot be explained currently may provide a new understanding of causal relationships. Third, the use of cluster-dropout models provided a powerful tool for distinguishing between individual genes and environmental factors that impact a specific physiological function from those that might affect all aspects of aging.

Many of the genes we identified are consistent hits in longevity GWAS analysis. Intuitively, this would be expected since aging is a risk factor for death. However, our model allows one to dig deeper and ask whether a longevity GWAS locus might be identified only because alleles prevent people from reaching an extremely old age. One could imagine that this is the case for

APOE, since *APOE4* individuals generally die prematurely of Alzheimer's disease (Olichney et al., 1997; Wright et al., 2019). However, we find that at all ages, even as young as 40 years, *APOE* genotype influences ΔAge (fig. 4f), perhaps due to its more general effects on lipid homeostasis (Abondio et al., 2019) or inflammation.

To gain a deeper understanding of the genetic signature of ΔAge, it might be prudent to consider genetic loci that have a strong association with ΔAge (say p-value<10^{-6}), even though they do not reach the threshold for genome-wide statistical significance. While some of the loci in this expanded list can be false positives, many of the potential genetic determinants identified this way are of potential interest. The full list of associations is available as supplementary summary statistics table.

When analyzing the many phenotypes that predict aging using PLS modelling, we discovered that only 9-11 axes are necessary to predict age. This suggests that there might be only ~10 independent systems (physiological networks) driving human aging. Interestingly, although overall the traits that figure most prominently into the sum of the principal components tend to map onto individual phenotypic clusters ("Dropout Clusters"), together the "meaning" of the differentially weighted sets that comprise each principal component is not obvious. For example, the 10 Dropout Clusters we used are not representative of the 10 axes identified in PLS analysis. It would be interesting to understand the physiological significance of each axis to better understand the process of aging. That said, the two most valuable phenotypes used in our study (those that had, overall, the most weight in age prediction) were forced vital capacity and blood pressure. Moreover, the genetic signature of ΔAge was similar to the genetic signature for FVC and blood pressure (figure 4e). These phenotypes are integrated, multidimensional health measurements. Using genetic information to better understand age-related phenotypes through PLS axis decomposition might be a fruitful direction for future research. Finally, from a practical perspective, we suggest that measuring human biological age using the 12 simple but diverse physiological measurements that together capture ~87% of the full ΔAge model (systolic blood pressure, forced expiratory volume, and so forth; see Fig. 2), might have actuarial and clinical value. For example, this physiological-age index could be measured longitudinally to learn how aging trajectories might be affected by environmental factors or anti-aging therapeutics.

Author Contributions

S.L. and C.K. conceived the idea for the study and its design, S.L. performed modelling and calculations, S.L. and C.K. wrote the manuscript, A.C. maintained and troubleshooted computing clusters and scripts necessary for data acquisition and GWAS.

Acknowledgments

Authors are grateful to the Calico community for support and discussions. Specifically, we are grateful to Eugene Melamud and his group for their support of the UKBB data framework, help with data processing and analysis, and numerous constructive discussions; we are grateful to Aarif Khakoo for his insights into disease-vs-aging paradigms; Kevin Wright and Graham Ruby for their discernment of human genetics of aging; and to David Botstein for his insights into statistical interpretation of our computational results and future applicability of such analysis. We are grateful to Madeleine Cule for supporting Calico's UKBB data interface and GWAS

cluster maintenance and for guidance with GWAS methodology. We are grateful to Amoolya Singh for providing guidance in regression modelling, statistical analysis and interpretation of the data. The authors are grateful to Jonathan K. Pritchard for constructive discussions about modelling, genetic analysis, and results interpretation. This study was carried out using UK Biobank Application number 44584, and we thank the participants in the UK Biobank study. This study was funded by Calico Life Sciences LLC.

Declaration of Interests

All coauthors work for Calico Life Sciences LLC, a pharmaceutical company engaged in understanding the biology of aging and development of therapies that ameliorate age-associated disease.

References

Abdellaoui A, Hugh-Jones D, Yengo L, Kemper KE, Nivard MG, Veul L, Holtz Y, Zietsch BP, Frayling TM, Wray NR, Yang J, Verweij KJH, Visscher PM. n.d. Genetic correlates of social stratification in Great Britain. doi:10.1038/s41562-019-0757-5

Abel E V., Goto M, Magnuson B, Abraham S, Ramanathan N, Hotaling E, Alaniz AA, Kumar-Sinha C, Dziubinski ML, Urs S, Wang L, Shi J, Waghray M, Ljungman M, Crawford HC, Simeone DM. 2018. HNF1A is a novel oncogene that regulates human pancreatic cancer stem cell properties. *Elife* **7**. doi:10.7554/eLife.33947

Abondio P, Sazzini M, Garagnani P, Boattini A, Monti D, Franceschi C, Luiselli D, Giuliani C. 2019. The genetic variability of APOE in different human populations and its implications for longevity. *Genes (Basel)*. doi:10.3390/genes10030222

Apfeld J, Kenyon C. 1999. Regulation of lifespan by sensory perception in Caenorhabditis elegans. *Nature* **402**:804–809. doi:10.1038/45544

Bae H, Gurinovich A, Karagiannis TT, Song Z, Leshchyk A, Li M, Andersen SL, Arbeev K, Yashin A, Zmuda J, An P, Feitosa M, Giuliani C, Franceschi C, Garagnani P, Mengel-From J, Atzmon G, Barzilai N, Puca A, Schork NJ, Perls TT, Sebastiani P. 2022. A Genome-Wide Association Study of 2304 Extreme Longevity Cases Identifies Novel Longevity Variants. *Int J Mol Sci* **24**. doi:10.3390/IJMS24010116

Chen Weiyang, Qian W, Wu G, Chen Weizhong, Xian B, Chen X, Cao Y, Green CD, Zhao F, Tang K, Han JDJ. 2015. Three-dimensional human facial morphologies as robust aging markers. *Cell Res* **25**:574–587. doi:10.1038/cr.2015.36

Chen Z, Boehnke M, Wen X, Mukherjee B. 2021. Revisiting the genome-wide significance threshold for common variant GWAS. *G3 Genes|Genomes|Genetics* **11**. doi:10.1093/G3JOURNAL/JKAA056

Chen Z, Raj A, Prateek G V., Di Francesco A, Liu J, Keyes BE, Kolumam G, Jojic V, Freund A. 2022. Automated, high-dimensional evaluation of physiological aging and resilience in outbred mice. *Elife* **11**. doi:10.7554/ELIFE.72664

Chetty R, Stepner M, Abraham S, Lin S, Scuderi B, Turner N, Bergeron A, Cutler D. 2016. The association between income and life expectancy in the United States, 2001-2014. *JAMA - Journal of the American Medical Association* **315**:1750–1766. doi:10.1001/jama.2016.4226

Coenen L, Lehallier B, Vries HE de, Middeldorp J. 2023. Markers of aging: Unsupervised integrated analyses of the human plasma proteome. *Frontiers in Aging* **4**. doi:10.3389/FRAGI.2023.1112109

Deelen J, Kettunen J, Fischer K, van der Spek A, Trompet S, Kastenmüller G, Boyd A, Zierer J, van den Akker EB, Ala-Korpela M, Amin N, Demirkan A, Ghanbari M, van Heemst D, Ikram MA, van Klinken JB, Mooijaart SP, Peters A, Salomaa V, Sattar N, Spector TD, Tiemeier H, Verhoeven A, Waldenberger M, Würtz P, Davey Smith G, Metspalu A, Perola M, Menni C, Geleijnse JM, Drenos F, Beekman M, Jukema JW, van Duijn CM, Slagboom PE. 2019. A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nat Commun* **10**:1–8. doi:10.1038/s41467-019-11311-9

Garratt M, Erturk I, Alonzo R, Zufall F, Leinders-Zufall T, Pletcher SD, Miller RA. 2022. Lifespan extension in female mice by early, transient exposure to adult female olfactory cues. *Elife* **11**. doi:10.7554/ELIFE.84060

Gibson J, Russ TC, Clarke TK, Howard DM, Hillary RF, Evans KL, Walker RM, Bermingham ML, Morris SW, Campbell A, Hayward C, Murray AD, Porteous DJ, Horvath S, Lu AT, McIntosh AM, Whalley HC, Marioni RE. 2019. A meta-analysis of genome-wide association studies of epigenetic age acceleration. *PLoS Genet* **15**:e1008104. doi:10.1371/JOURNAL.PGEN.1008104

Graham Ruby J, Wright KM, Rand KA, Kermany A, Noto K, Curtis D, Varner N, Garrigan D, Slinkov D, Dorfman I, Granka JM, Byrnes J, Myres N, Ball C. 2018. Estimates of the heritability of human longevity are substantially inflated due to assortative mating. *Genetics* **210**:1109–1124. doi:10.1534/genetics.118.301613

Hanlon Id P, Jani BD, Nicholl Id B, Lewsey J, Mcallister Id DA, Mair Id FS. 2022. Associations between multimorbidity and adverse health outcomes in UK Biobank and the SAIL Databank: A comparison of longitudinal cohort studies. doi:10.1371/journal.pmed.1003931

Haworth S, Mitchell R, Corbin L, Wade KH, Dudding T, Budu-Aggrey A, Carslake D, Hemani G, Paternoster L, Smith GD, Davies N, Lawson DJ, J. Timpson N. 2019. Apparent latent structure within the UK Biobank sample has implications for epidemiological analysis. *Nat Commun* **10**. doi:10.1038/S41467-018-08219-1

Jones M, Denieffe S, Griffin C, Tinago W, Fitzgibbon MC. 2017. Evaluation of cystatin C in malignancy and comparability of estimates of GFR in oncology patients. *Pract Lab Med* **8**:95–104. doi:10.1016/j.plabm.2017.05.005

Joshi PK. n.d. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity OPEN. doi:10.1038/s41467-017-00934-5

Justice JN, Ferrucci L, Newman AB, Aroda VR, Bahnson JL, Divers J, Espeland MA, Marcovina S, Pollak MN, Kritchevsky SB, Barzilai N, Kuchel GA. 2018. A framework for selection of blood-based biomarkers for geroscience-guided clinical trials: report from the TAME Biomarkers Workgroup. *Geroscience* **40**:419–436. doi:10.1007/S11357-018-0042-Y

Kaur G, Levy E. 2012. Cystatin C in Alzheimer's disease. *Front Mol Neurosci*. doi:10.3389/fnmol.2012.00079

Kirkwood TBL. 2015. Deciphering death: a commentary on Gompertz (1825) 'On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies.' *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**:20140379. doi:10.1098/rstb.2014.0379

Li LB, Lei H, Arey RN, Li P, Liu J, Murphy CT, Xu XZS, Shen K. 2016. The neuronal kinesin UNC-104/KIF1A is a key regulator of synaptic aging and insulin signaling-regulated memory. *Curr Biol* **26**:605. doi:10.1016/J.CUB.2015.12.068

Libert S, Zwiener J, Chu X, VanVoorhies W, Roman G, Pletcher SD. 2007. Regulation of Drosophila life span by olfaction and food-derived odors. *Science* **315**:1133–1137. doi:10.1126/SCIENCE.1136610

Lu AT, Xue L, Salfati EL, Chen BH, Ferrucci L, Levy D, Joehanes R, Murabito JM, Kiel DP, Tsai PC, Yet I, Bell JT, Mangino M, Tanaka T, McRae AF, Marioni RE, Visscher PM, Wray NR, Deary IJ, Levine ME, Quach A, Assimes T, Tsao PS, Absher D, Stewart JD, Li Y, Reiner AP, Hou L, Baccarelli AA, Whitsel EA, Aviv A, Cardona A, Day FR, Wareham NJ, Perry JRB, Ong KK, Raj K, Lunetta KL, Horvath S. 2018. GWAS of epigenetic aging rates in blood reveals a critical role for TERT. *Nature Communications 2018 9:1* **9**:1–13. doi:10.1038/s41467-017-02697-5

Marioni RE, Shah S, McRae AF, Chen BH, Colicino E, Harris SE, Gibson J, Henders AK, Redmond P, Cox SR, Pattie A, Corley J, Murphy L, Martin NG, Montgomery GW, Feinberg AP, Fallin MD, Multhaup ML, Jaffe AE, Joehanes R, Schwartz J, Just AC, Lunetta KL, Murabito JM, Starr JM, Horvath S, Baccarelli AA, Levy D, Visscher PM, Wray NR, Deary IJ. 2015. DNA methylation age of blood predicts all-cause mortality in later life. *Genome Biol* **16**:25. doi:10.1186/s13059-015-0584-6

McCartney DL, Min JL, Richmond RC, Lu AT, Sobczyk MK, Davies G, Broer L, Guo X, Jeong A, Jung J, Kasela S, Katrinli S, Kuo PL, Matias-Garcia PR, Mishra PP, Nygaard M, Palviainen T, Patki A, Raffield LM, Ratliff SM, Richardson TG, Robinson O, Soerensen M, Sun D, Tsai PC, van der Zee MD, Walker RM, Wang X, Wang Y, Xia R, Xu Z, Yao J, Zhao W, Correa A, Boerwinkle E, Dugué PA, Durda P, Elliott HR, Gieger C, de Geus EJC, Harris SE, Hemani G, Imboden M, Kähönen M, Kardia SLR, Kresovich JK, Li S, Lunetta KL, Mangino M, Mason D, McIntosh AM, Mengel-From J, Moore AZ, Murabito JM, Ollikainen M, Pankow JS, Pedersen NL, Peters A, Polidoro S, Porteous DJ, Raitakari O, Rich SS, Sandler DP, Sillanpää E, Smith AK, Southey MC, Strauch K, Tiwari H, Tanaka T, Tillin T, Uitterlinden AG, Van Den Berg DJ, van Dongen J, Wilson JG, Wright J, Yet I, Arnett D, Bandinelli S, Bell JT, Binder AM, Boomsma DI, Chen W, Christensen K, Conneely KN, Elliott P, Ferrucci L, Fornage M, Hägg S, Hayward C, Irvin M, Kaprio J, Lawlor DA, Lehtimäki T, Lohoff FW, Milani L, Milne RL, Probst-Hensch N, Reiner AP, Ritz B, Rotter JI, Smith JA, Taylor JA, van Meurs JBJ, Vineis P, Waldenberger M, Deary IJ, Relton CL, Horvath S, Marioni RE. 2021. Genome-wide association studies identify 137 genetic loci for DNA methylation biomarkers of aging. *Genome Biol* **22**:1–25. doi:10.1186/S13059-021-02398-9/FIGURES/4

Melzer D, Pilling LC, Ferrucci L. 2020. The genetics of human ageing. *Nat Rev Genet*. doi:10.1038/s41576-019-0183-6

Milman S, Barzilai N. 2016. Dissecting the mechanisms underlying unusually successful human health span and life span. *Cold Spring Harb Perspect Med* **6**. doi:10.1101/cshperspect.a025098

Min KJ, Lee CK, Park HN. 2012. The lifespan of Korean eunuchs. *Current Biology*. doi:10.1016/j.cub.2012.06.036

Olichney JM, Sabbagh MN, Hofstetter CR, Galasko D, Grundman M, Katzman R, Thal LJ. 1997. The impact of apolipoprotein E4 on cause of death in Alzheimer's disease. *Neurology* **49**:76–81. doi:10.1212/WNL.49.1.76

Pilling LC, Kuo CL, Sicinski K, Tamosauskaite J, Kuchel GA, Harries LW, Herd P, Wallace R, Ferrucci L, Melzer D. 2017. Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging* **9**:2504–2520. doi:10.18632/aging.101334

Qiu W, Chen H, Kaeberlein M, Lee S-I, Allen PG. 2022. An explainable AI framework for interpretable biological age. *medRxiv* 2022.10.05.22280735. doi:10.1101/2022.10.05.22280735

Schächter F, Faure-Delanef L, Guénot F, Rouger H, Froguel P, Lesueur-Ginot L, Cohen D. 1994. Genetic associations with human longevity at the APOE and ACE loci. *Nature Genetics 1994 6:1* **6**:29–32. doi:10.1038/ng0194-29

Sebastiani P, Gurinovich A, Nygaard M, Sasaki T, Sweigart B, Bae H, Andersen SL, Villa F, Atzmon G, Christensen K, Arai Y, Barzilai N, Puca A, Christiansen L, Hirose N, Perls TT. 2019. APOE Alleles and Extreme Human Longevity. *The Journals of Gerontology: Series A* **74**:44–51. doi:10.1093/GERONA/GLY174

Shen S, Li C, Xiao L, Wang X, Lv H, Shi Y, Li Y, Huang Q. 2020. Whole-genome sequencing of Chinese centenarians reveals important genetic variants in aging WGS of centenarian for genetic analysis of aging. *Hum Genomics* **14**. doi:10.1186/S40246-020-00271-7

Shepherd M, Shields B, Ellard S, Rubio-Cabezas O, Hattersley AT. 2009. A genetic diagnosis of HNF1A diabetes alters treatment and improves glycaemic control in the majority of insulin-treated patients. *Diabetic Medicine* **26**:437–441. doi:10.1111/j.1464-5491.2009.02690.x

Tian YE, Cropley V, Maier AB, Lautenschlager NT, Breakspear M, Zalesky A. 2023. Heterogeneous aging across multiple organ systems and prediction of chronic disease and mortality. *Nature Medicine 2023 29:5* **29**:1221–1231. doi:10.1038/s41591-023-02296-6

Timmers PRHJ, Mounier N, Lall K, Fischer K, Ning Z, Feng X, Bretherick AD, Clark DW, Shen X, Esko T, Kutalik Z, Wilson JF, Joshi PK. 2019. Genomics of 1 million parent lifespans implicates novel pathways and common diseases and distinguishes survival chances. *Elife* **8**:1–40. doi:10.7554/ELIFE.39856

Timmers PRHJ, Tiys ES, Sakaue S, Akiyama M, Kiiskinen TTJ, Zhou W, Hwang S-J, Yao C, Kamatani Y, Zhou W, Deelen J, Levy D, Ganna A, Kamatani Y, Okada Y, Joshi PK, Wilson JF, Tsepilov YA. 2022. Mendelian randomization of genetically independent aging phenotypes identifies LPA and VCAM1 as biological targets for human aging. *Nature Aging 2022 2:1* **2**:19–30. doi:10.1038/s43587-021-00159-8

Timmers PRHJ, Wilson JF, Joshi PK, Deelen J. 2020. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nature Communications 2020 11:1* **11**:1–10. doi:10.1038/s41467-020-17312-3

van der Laan SW, Fall T, Soumaré A, Teumer A, Sedaghat S, Baumert J, Zabaneh D, van Setten J, Isgum I, Galesloot TE, Arpegård J, Amouyel P, Trompet S, Waldenberger M, Dörr M, Magnusson PK, Giedraitis V, Larsson A, Morris AP, Felix JF, Morrison AC, Franceschini N, Bis JC, Kavousi M, O'Donnell C, Drenos F, Tragante V, Munroe PB, Malik R, Dichgans M, Worrall BB, Erdmann J, Nelson CP, Samani NJ, Schunkert H, Marchini J, Patel RS, Hingorani AD, Lind L, Pedersen NL, de Graaf J, Kiemeney LALM, Baumeister SE, Franco OH, Hofman A, Uitterlinden AG, Koenig W, Meisinger C, Peters A, Thorand B, Jukema JW, Eriksen BO, Toft I, Wilsgaard T, Onland-Moret NC, van der Schouw YT, Debette S, Kumari M, Svensson P, van der Harst P, Kivimaki M, Keating BJ, Sattar N, Dehghan A, Reiner AP, Ingelsson E, den Ruijter HM, de Bakker PIW, Pasterkamp G, Ärnlöv J, Holmes M V., Asselbergs FW. 2016. Cystatin C and Cardiovascular Disease: A Mendelian Randomization Study. *J Am Coll Cardiol* **68**:934–945. doi:10.1016/j.jacc.2016.05.092

Wang J, Geng L. 2019. Effects of socioeconomic status on physical and psychological health: Lifestyle as a mediator. *Int J Environ Res Public Health* **16**. doi:10.3390/ijerph16020281

Wilmanski T, Diener C, Rappaport N, Patwardhan S, Wiedrick J, Lapidus J, Earls JC, Zimmer A, Glusman G, Robinson M, Yurkovich JT, Kado DM, Cauley JA, Zmuda J, Lane NE, Magis AT, Lovejoy JC, Hood L, Gibbons SM, Orwoll ES, Price ND. 2021. Gut microbiome pattern reflects healthy aging and predicts survival in humans. *Nat Metab* **3**:274. doi:10.1038/S42255-021-00348-0

Wright KM, Rand KA, Kermany A, Noto K, Curtis D, Garrigan D, Slinkov D, Dorfman I, Granka JM, Byrnes J, Myres N, Ball CA, Ruby JG. 2019. A prospective analysis of genetic variants associated with human lifespan. *G3: Genes, Genomes, Genetics* **9**:2863–2878. doi:10.1534/g3.119.400448

Xia X, Chen X, Wu G, Li F, Wang Y, Chen Y, Chen M, Wang X, Chen Weiyang, Xian B, Chen Weizhong, Cao Y, Xu C, Gong W, Chen G, Cai D, Wei W, Yan Y, Liu K, Qiao N, Zhao X, Jia J, Wang W, Kennedy BK, Zhang K, Cannistraci C V., Zhou Y, Han JDJ. 2020. Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. *Nat Metab* **2**:946–957. doi:10.1038/S42255-020-00270-X

Xu M, Zhao J, Zhang Y, Ma X, Dai Q, Zhi H, Wang B, Wang L. 2016. Apolipoprotein E Gene Variants and Risk of Coronary Heart Disease: A Meta-Analysis. *Biomed Res Int* **2016**. doi:10.1155/2016/3912175

Yang F, Sun L, Zhu X, Han J, Zeng Y, Nie C, Yuan H, Li X, Shi X, Yang Y, Hu C, Lv Z, Huang Z, Zheng C, Liang S, Huang J, Wan G, Qi K, Qin B, Cao S, Zhao X, Zhang Y, Yang Z. 2017. Identification of new genetic variants of HLA-DQB1 associated with human longevity and lipid homeostasis—a cross-sectional study in a Chinese population. *Aging (Albany NY)* **9**:2316. doi:10.18632/AGING.101323

Zenin A, Tsepilov Y, Sharapov S, Getmantsev E, Menshikov LI, Fedichev PO, Aulchenko Y. n.d. Identification of 12 genetic loci associated with human healthspan. *Akademika Kurchatova pl* **10**. doi:10.1038/s42003-019-0290-0

Zheng J, Mesut Erzurumluoglu A, Elsworth BL, Kemp JP, Howe L, Haycock PC, Hemani G, Tansey K, Laurin C, Warrington NM, Finucane HK, Price AL, Bulik-Sullivan BK, Anttila V, Paternoster L, Gaunt TR, Evans DM, Neale BM. 2017. Databases and ontologies LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**:272–279. doi:10.1093/bioinformatics/btw613

Zullo JM, Drake D, Aron L, O'Hern P, Dhamne SC, Davidsohn N, Mao CA, Klein WH, Rotenberg A, Bennett DA, Church GM, Colaiácovo MP, Yankner BA. 2019. Regulation of lifespan by neural excitation and REST. *Nature* **574**:359–364. doi:10.1038/S41586-019-1647-8